Determining the exponent b using Linear Regression

To determine the exponent b and the activities of the ^{238}U and ^{232}Th series, one can use an unweighted linear regression formula or perform a parameter search to minimize a chi-square funtion. Below, we derive the formula for we use for the linear regression method with the data from our soil sample.

For the linear regression approach, one starts with the equations:

$$
\begin{aligned}
(C_1/Y)_i &= A_I \epsilon_{1460} (E_{1i}/1460)^b \qquad (i = 1, 2, \cdots, N_I) \\
(C_2/Y)_i &= A_{II} \epsilon_{1460} (E_{2i}/1460)^b \qquad (i = 1, 2, \cdots, N_{II})
\end{aligned}
$$

where A_I is the activity for decay series I (i.e. ^{238}U, A_{II} is the activity for the decay series II (i.e. ^{232}Th) and the exponent b is the same for both the sources. We need to vary A_I, A_{II} and b for a "best fit" to the data.

To linearize the equations, one takes the log of both sides of the equations:

$$
\begin{aligned}
ln((C_1/Y)_i) &= b\, ln(E_{1i}/1460) + ln(A_I \epsilon_{1460}) \qquad (i = 1 \to N_I) \\
ln((C_2/Y)_i) &= b\, ln(E_{2i}/1460) + ln(A_{II} \epsilon_{1460}) \qquad (i = 1 \to N_{II})
\end{aligned}
$$

We define $y_{1i} \equiv ln((C_1/Y)_i)$ and $y_{2i} \equiv ln((C_2/Y)_i)$ for the count data, $x_{1i} \equiv ln(E_{1i}/1460)$ and $x_{2i} \equiv ln(E_{2i}/1460)$ for the energy data, $k_1 \equiv ln(A_I \epsilon_{1460})$, $k_2 \equiv ln(A_{II})\epsilon_{1460}$, $n1 \equiv N_I$ and $n2 \equiv N_{II}$. Then the equations become:

$$
\begin{aligned}
y_{1i} &= b\, x_{1i} + k_1 \qquad (i = 1 \to n1) \\
y_{2i} &= b\, x_{2i} + k_2 \qquad (i = 1 \to n2)
\end{aligned}
$$

There are 3 free parameters to vary to best fit the data: b, k_1, and k_2. To determine the "best fit" values we use linear regression. The chi-square function χ^2 is defined as:

$$
\chi^2 \equiv \sum_{i=1}^{n1} (b\, x_{1i} + k_1 - y_{1i})^2 + \sum_{i=1}^{n2} (b\, x_{2i} + k_2 - y_{2i})^2 \qquad (1)
$$

The "best fit" values are the ones that minimize the χ^2 function. At the minimum value of χ^2, the derivatives with respect to each of the free parameters are zero: $\frac{\partial \chi^2}{\partial b} = 0$, $\frac{\partial \chi^2}{\partial k_1} = 0$, and $\frac{\partial \chi^2}{\partial k_2} = 0$.

The derivative with respect to b yields:

$$\frac{\partial\chi^2}{\partial b} = \sum_{i=1}^{n1} 2(b\,x_{1i} + k_1 - y_{1i})x_{1i} + \sum_{i=1}^{n2} 2(b\,x_{2i} + k_2 - y_{2i})x_{2i} = 0 \qquad (2)$$

The above equation can be written as

$$bX^2 + k_1X_1 + k_2X_2 - \sum_{i=1}^{n1} y_{1i}x_{1i} - \sum_{i=1}^{n2} y_{2i}x_{2i} = 0 \qquad (3)$$

where $X^2 \equiv \sum_{i=1}^{n1} x_{1i}^2 + \sum_{i=1}^{n2} x_{2i}^2$, $X_1 \equiv \sum_{i=1}^{n1} x_{1i}$, and $X_2 \equiv \sum_{i=1}^{n2} x_{2i}$. The derivative with respect to k_1 yields:

$$\begin{aligned} \frac{\partial\chi^2}{\partial k_1} &= \sum_{i=1}^{n1} 2(b\,x_{1i} + k_1 - y_{1i}) = 0 && (4) \\ bX_1 + (n1)k_1 - Y_1 &= 0 && (5) \end{aligned}$$

where $Y_1 \equiv \sum_{i=1}^{n1} y_{1i}$. Similarily, for the last parameter we have

$$bX_2 + (n2)k_2 - Y_2 = 0 \qquad (6)$$

with $Y_2 \equiv \sum_{i=1}^{n2} y_{2i}$.
Combining equations 3, 5, and 6, we obtain one equation involving only b:

$$b(X^2 - \frac{X_1^2}{n1} - \frac{X_2^2}{n2}) = \sum_{i=1}^{n1} x_{1i}y_{1i} + \sum_{i=1}^{n2} x_{2i}y_{2i}) - \frac{Y_1X_1}{n1} - \frac{Y_2X_2}{n2} \qquad (7)$$

Which can be solved for b

$$b = \frac{\sum_{i=1}^{n1} x_{1i}y_{1i} + \sum_{i=1}^{n2} x_{2i}y_{2i} - Y_1X_1/n1 - Y_2X_2/n2}{X^2 - (X_1)^2/(n1) - (X_2)^2/(n1)} \qquad (8)$$

Once b is determined, one can solve for the other two parameters:

$$\begin{aligned} k_1 &= (Y_1 - bX_1)/(n1) \\ k_2 &= (Y_2 - bX_2)/(n2) \end{aligned}$$

and then the values of $A_I\epsilon_{1460}$ and $A_{II}\epsilon_{1460}$ can be determined.. The formulas above are a minor generalization of the linear regression formula commonly used in undergraduate physics labs.